RESEARCH ARTICLE

Methods in Ecology and Evolution | BRITISH ECOLOGICAL SOCIETY

Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity

Wilson J. Wright[1] (iD) | Kathryn M. Irvine[2] (iD) | Emily S. Almberg[3] | Andrea R. Litt[1] (iD)

[1]Department of Ecology, Montana State University, Bozeman, MT, USA

[2]U.S. Geological Survey, Northern Rocky Mountain Science Center, Bozeman, MT, USA

[3]Montana Fish, Wildlife & Parks, Wildlife Division, Bozeman, MT, USA

Correspondence
Wilson J. Wright
Email: wilson.wright@msu.montana.edu

Abstract

1. Surveying wildlife communities provides data for informing conservation and management decisions that affect multiple species. Autonomous recording units (ARUs) can efficiently gather community data for a variety of taxa, but generally require software algorithms to classify each recorded call to a species. Species classification errors are possible during this process and result in both false-negative and false-positive detections. Available approaches for analysing ARU data do not model the species classification probabilities, meaning erroneous detections are attributed to an omnibus source instead of the presence of another species. Additionally, counts of call recordings for each species are often summarized to binary detection data for analyses. Expanding statistical models to capture these nuances of ARU data would allow for improved inferences about occupancy and relative activity.

2. Motivated by bat acoustic surveys, we developed a model to analyse counts of call recordings from multiple species simultaneously while accounting for species classification errors. Our model expands on previously developed false-positive occupancy models to better describe acoustic data. We used simulations to compare our model to other false-positive occupancy models for an example scenario with ARU data from two species. We also analyse acoustic data for eight bat species in Montana using our model.

3. In simulations, single-species models resulted in biased estimates of occupancy and relative activity because they failed to associate false positives with the presence of the second species. Models analysing binary observations ignored available information on relative activity and led to less precise estimates. Applying our model to bat acoustic data from Montana allowed for species-specific estimates of occupancy and relative activity. This analysis illustrates the flexibility in our model framework while also highlighting the assumptions and data requirements for implementation. Specifically, additional information on the species classification probabilities is needed and we discuss considerations for reliably estimating these parameters.

4. Directly modelling the species classification probabilities allows for improved ecological inferences for both occupancy and relative activity using community ARU data. Our statistical framework helps address the challenges posed by acoustic data, allowing ecologists to better utilize this technology to monitor wildlife communities.

1 | INTRODUCTION

Surveying wildlife communities provides data for informing conservation and management decisions that affect multiple species. Autonomous recording units (ARUs) efficiently gather community data by passively recording animal vocalizations (Gibb, Browning, Glover-Kapfer, & Jones, 2019), generally for multiple time periods ('visits') at each surveyed location ('site'). These data, including counts of call recordings and corresponding species classifications, can be used to investigate various ecological questions and are applicable for surveying multiple taxa (e.g. anurans, bats, birds). However, due to the large volumes of data typically collected, most studies using acoustic surveys require classification software to identify the species of each call recording (Gibb et al., 2019). This automated process includes species classification errors that lead to both false-negative and false-positive detections. For instance, when a species is present, false-negative detections can result from successfully recording its calls but misclassifying them as alternative species. These errors are in addition to false negatives from failing to record any of its calls. False-positive detections at sites where a species is absent are often due to misclassifying recorded calls from another species. Estimating the ecological parameters of interest, while addressing these errors is an important consideration when analysing ARU data.

Occupancy models (MacKenzie et al., 2002) are a natural framework for analysing ARU data when visits are summarized to detection/non-detection observations for each species (e.g. Banner et al., 2018; Rodhouse et al., 2019). Originally developed to account for false negatives, standard occupancy models assume that all false positives are removed (MacKenzie et al., 2002). Completely eliminating false positives from ARU data is generally cost prohibitive because it requires manually confirming at least one recording for every visit. False positives are an important source of errors in many wildlife surveys (Chambert, Miller, & Nichols, 2015; Guillera-Arroita, Lahoz-Monfort, van Rooyen, Weeks, & Tingley, 2017; Miller et al., 2011) and can negatively impact inferences when ignored, even when infrequent (Miller et al., 2015). Consequently, expanded occupancy models that account for false positives were developed and have been applied to ARU data (Banner et al., 2018; Chambert, Waddle, Miller, Walls, & Nichols, 2018). Other approaches for ARU data utilize the degree of certainty generated by automated classifications to limit false positives and analyse the resulting data with false-positive occupancy models (Balantic & Donovan, 2019) or other methods (Barré et al., 2019). However, existing statistical models for ARU data are unable to attribute erroneous detections to the presence of another species while also analysing the counts of call recordings from each visit to a site.

False-positive occupancy models, including those applied to ARU data, typically estimate parameters for one species (or species group) and attribute erroneous detections to an omnibus source (Balantic & Donovan, 2019; Banner et al., 2018; Chambert et al., 2015; Chambert, Waddle, et al., 2018; Miller et al., 2011). However, Chambert, Campbell Grant, et al. (2018) showed single-species models are inadequate when false positives result from misclassifying a second species. Acoustic surveys of more than two species compound this issue because recorded calls may be classified to one of many possible species with similar acoustic characteristics. These classification errors mean ARU data inherently include multiple species and analyses should reflect this characteristic when modelling false positives. Another characteristic of ARU data is detection counts are available for each visit to a site and allow for ecological inferences beyond species occurrence. For instance, anuran surveys collect calling rates that are used as an index for abundance (Royle, 2004; Royle & Link, 2005) and bat acoustic surveys monitor activity using the number of recorded call files (Johnson, Gates, & Ford, 2008; Nocera, Ford, Silvis, & Dobony, 2019). Estimates of relative activity are useful for addressing questions at smaller scales and provide a more sensitive metric for monitoring populations. When false positives occur, recording counts can also better inform the probability a site is occupied (Chambert, Waddle, et al., 2018).

We develop a community occupancy model for analysing call recording counts while explicitly incorporating the species classification probabilities. Using the number of recorded calls classified to each species and the probabilities of classifying a recording to each potential species, we account for both false negatives and false positives resulting from the species classification process. By modelling multiple species simultaneously, our approach uses additional information from ARU surveys when estimating species-specific occupancy and relative activity. Our model extends false-positive occupancy models (Chambert, Campbell Grant, et al., 2018; Chambert et al., 2015; Chambert, Waddle, et al., 2018) to analyse call counts instead of binary observations and explicitly incorporates data from multiple species. We also use a simulation study to compare different false-positive occupancy models for an example scenario of collecting ARU data from two species.

We motivate our model using multi-species acoustic bat surveys and apply it to data collected in Montana, USA. Many bat species face multiple conservation threats, including the spread of white-nose syndrome (WNS), wind energy development, and habitat change (O'Shea, Cryan, Hayman, Plowright, & Streicker, 2016). Acoustic data provide valuable information for monitoring populations because bats are nocturnal and elusive, making them difficult to study using capture techniques (e.g. mist netting). Bat conservation research has focused on using acoustic data to map species distributions (e.g. Rodhouse et

FIGURE 1 Diagram depicting call detection and automated species classification for two species. Dashed and solid lines indicate incorrect and correct classifications, respectively. With automated classification, the observed call file counts for species 1 can include call files originating from both species
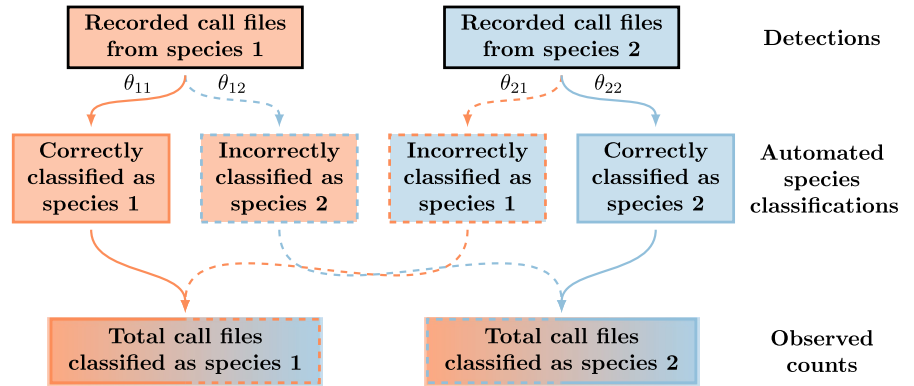
al., 2019) or model patterns in activity (e.g. Nocera et al., 2019). We introduce a comprehensive framework for modelling acoustic data which accounts for species classification errors while simultaneously estimating species-specific occupancy and relative activity.

2 | MATERIALS AND METHODS

2.1 | Bat acoustic data description

First, we provide a general description of bat acoustic data to motivate our model development. Stationary acoustic detectors for bats consist of ultrasonic microphones elevated from the ground using poles. Detectors are deployed at suitable locations (e.g. see Maxell, 2015) and record bat echolocation calls as individuals navigate throughout the area. After detectors are retrieved, acoustic files are processed by software to remove non-bat recordings (Parsons & Szewczak, 2009). This initial filtering process reliably discards, for the most part, all files without bat calls (Banner et al., 2018; Reichert et al., 2018). Then software classifies each call file to one bat species, if possible (see Banner et al., 2018). These automated species classifications include errors that we aim to model. For example, consider the scenario where ARU surveys are conducted within the ranges of two bat species with similar echolocation call characteristics. An observed count of recordings classified as species 1 can include a combination of correctly classified call files and incorrectly classified call files from species 2 (Figure 1). In this scenario, false-positive detections for species 1 are only possible at sites where species 2 is present. That is, the only source of false positives is incorrectly classified call files from the other species. False-negative detections of species 1 occur when no calls from species 1 are recorded or when every recorded call file is misclassified as species 2. This example shows how misclassifying observations results in false negatives for one species becoming the false positives for another.

The hidden cost of bat acoustic surveys is that experts visually examine characteristics of the sonograms (images depicting attributes of sound) to assess whether the classification software assigned the correct species. These manual species identifications are made by individuals with extensive training and considered error-free (e.g. Gibb et al., 2019). Manual identifications are used to determine the error rates for various automated classification programs. When estimating occupancy, one approach is to confirm at least one detection of the species per visit and analyse these data using standard single-species occupancy models (e.g. Banner et al., 2018). Alternatively, a combination of manually confirmed and unconfirmed calls can be analysed with false-positive occupancy models (Chambert et al., 2015), requiring manual confirmation of fewer calls (Banner et al., 2018). Modelling species-specific activity rates requires human verification of all recordings, reflecting a costly and often unrealistic expectation. We develop an approach that harnesses acoustic data from multiple species to simultaneously estimate occupancy and relative activity while modelling the species classification probabilities. This reduces the amount of manual confirmation needed. Our approach is applicable to multi-species datasets when the counts of call recordings include species classification errors.

2.2 | General model framework

Let $i = \{1, \dots, n\}$ index sites, $j = \{1, \dots, J\}$ index visits within each site, and $k = \{1, \dots, K\}$ index possible species surveyed during a study. For site i, species k is present (1) or not (0) with probability ψ_{ik} and we model this latent occupancy state as follows:

$$Z_{ik} \sim \text{Bernoulli}(\psi_{ik}). \tag{1}$$

We assume the occupancy states for each species are independent, but this could be modified to allow for species interactions. Site-level covariates modelling the probabilities of occupancy for each species can be incorporated using a generalized linear model framework with $g(\psi_{ik}) = \mathbf{X}_i \boldsymbol{\alpha}_k$ where $g()$ denotes an appropriate link function (e.g. logit, probit). We have represented a row-vector of covariates for site i as $\mathbf{X}_i$ and a vector of coefficients for species k as $\boldsymbol{\alpha}_k$.

For sites occupied by a species, consider the number of detections on each visit a Poisson distributed random variable. Given site i is occupied by species k, we model the detection count for visit j as

$$[Y_{ijk} | Z_{ik} = 1] \sim \text{Poisson}(\lambda_{ijk}), \tag{2}$$

where λ_{ijk} is the expected number of detections per visit. In this formulation, false negatives can result from the species going undetected during visits to occupied sites as a result of $Y_{ijk} = 0$ from the associated Poisson distribution. Covariates can be included to model the detection rate in a generalized linear model framework where $\log(\lambda_{ijk}) = \mathbf{V}_{ij}\boldsymbol{\beta}_k$. Let $\mathbf{V}_{ij}$ denote a row-vector combining site-level and visit-level covariates, $\boldsymbol{\beta}_k$ denote the vector of associated coefficients for species k, and $\log()$ represent the natural logarithm link function. So far, we have described detections for a visit without considering the species classification probabilities. If the species associated with each detection is known without error, the detection counts for each species (Y_{ijk}) are directly observable with Equations 1 and 2 fully describing the observed data. In reality, however, these counts are not directly observable because individual detections may be incorrectly classified (Figure 1).

Let $\theta_{kk'}$ be the probability that a single detection of species k is classified as species k' (e.g. Figure 1 shows an example with $K = 2$ species). This includes the probability of a correct species classification when $k' = k$. Across all possible species that could be classified ($k' = \{1, \ldots, K\}$), these probabilities sum to one. For detections of species k, let $\boldsymbol{\Theta}_k = [\theta_{k1}, \ldots, \theta_{kK}]^\top$ be a vector containing the probabilities of classifying a detection as each possible species. The entire confusion matrix, $\boldsymbol{\Theta}$, is defined using the probability vectors from every species so that

$$\boldsymbol{\Theta} = \begin{bmatrix} \boldsymbol{\Theta}_1^\top \\ \boldsymbol{\Theta}_2^\top \\ \vdots \\ \boldsymbol{\Theta}_K^\top \end{bmatrix} = \begin{bmatrix} \theta_{11} & \theta_{12} & \cdots & \theta_{1K} \\ \theta_{21} & \theta_{22} & \cdots & \theta_{2K} \\ \vdots & \vdots & \ddots & \vdots \\ \theta_{K1} & \theta_{K2} & \cdots & \theta_{KK} \end{bmatrix}$$

with rows corresponding to the actual recorded species and columns indicating the species assigned from the classification algorithm. The diagonal elements of $\boldsymbol{\Theta}$ correspond to the probabilities of a correct species classification. Given the number of detections for species k, y_{ijk}, the number classified to each species for site i and visit j is a multinomial random variable with probabilities $\boldsymbol{\Theta}_k$. That is,

$$[\mathbf{C}_{ijk} \mid Z_{ik} = 1, Y_{ijk} = y_{ijk}] \sim \text{Multinomial}(y_{ijk}, \boldsymbol{\Theta}_k), \tag{3}$$

where $\mathbf{C}_{ijk}$ is the K-length vector containing the number of detections from species k classified to each possible species.

Let $C_{ijkk'}$ denote the components of $\mathbf{C}_{ijk}$ and represent the number of detections from species k classified as species k' for visit j to site i. Again, because the true species identities of individual detections are unknown, these counts are unobservable. Instead, the observed data consist of the total number of detections classified as species k', which we denote as

$$C_{ij \cdot k'} = \sum_{k=1}^{K} c_{ijkk'} \tag{4}$$

for site i and visit j. The detections included in this sum are ambiguous because they potentially include detections from any of the K possible species. This component of the model appropriately adjusts the estimates of occupancy and relative activity to account for the uncertainty in the species classifications. The representation in Equations 1–4 most directly describes each component of this model, but an alternative representation (see Supporting Information S1) is useful computationally and for comparisons with other models.

As with other false-positive occupancy models, this model is only identifiable with additional information about the classification process (Chambert et al., 2015; Guillera-Arroita et al., 2017). Specifically, information is needed to estimate the classification probabilities ($\theta_{kk'}$) separately from the detection rates (λ_{ijk}). This can include strong prior information (using Bayesian methods), auxiliary data in the form of automated species classifications for call recordings of known species, or using error-free methods to confirm the species classifications for a subset of the observed data. These options are analogous to those used with other false-positive occupancy models (Chambert et al., 2015; Guillera-Arroita et al., 2017). For our approach, auxiliary data inform the classification probabilities through a multinomial probability model. The species classification probabilities can be estimated using recordings collected in the field if an error-free classification method is available. Consider the case where a subset of the sites ('confirmed sites') have all recordings from every visit identified using the error-free method. Data from confirmed sites are incorporated into an analysis with Equations 1–3 because the actual species for each detection are known ($\mathbf{C}_{ijk}$ observable). This model structure could be modified if only a portion of visits to a site are confirmed (similar to Banner et al., 2018).

2.3 | Comparisons to other false-positive occupancy models

Our model generalizes other false-positive occupancy models by analysing detection counts from multiple species while modelling the species classification probabilities. In a two species scenario ($K = 2$), our model is related to other false-positive occupancy models (Chambert, Campbell Grant, et al., 2018; Chambert et al., 2015; Chambert, Waddle, et al., 2018) after slight modifications. Consider the two-species scenario where the primary interest is making inferences for a focal species. The second species still needs consideration because its call files can be misclassified as the focal species, leading to false positives. Similarly, false negatives can result when call files from the focal species are erroneously classified as the second species. For our general model, the observed data include count detections for both species (Supporting Information S2.1). Alternative models analyse detections summarized to binary indicators (i.e. let $D_{ij \cdot k'} = I(C_{ij \cdot k'} > 0)$) and/or only data from the focal species (Figure 2). Specifically, we consider the following alternative false-positive models: two-species with binary detections (Supporting Information S2.2; Chambert, Campbell Grant, et al., 2018), single-species with count detections (Supporting Information S2.3; Chambert, Waddle, et al., 2018), or single-species with binary detections (Supporting Information S2.4; Chambert et

FIGURE 2 When ARU surveys conflate two species, our model (two-species, count detections) is related to other false-positive occupancy models (Chambert, Campbell Grant, et al., 2018; Chambert et al., 2015; Chambert, Waddle, et al., 2018). These alternative approaches rely on simplifying assumptions to analyse binary observations or only data for a focal species
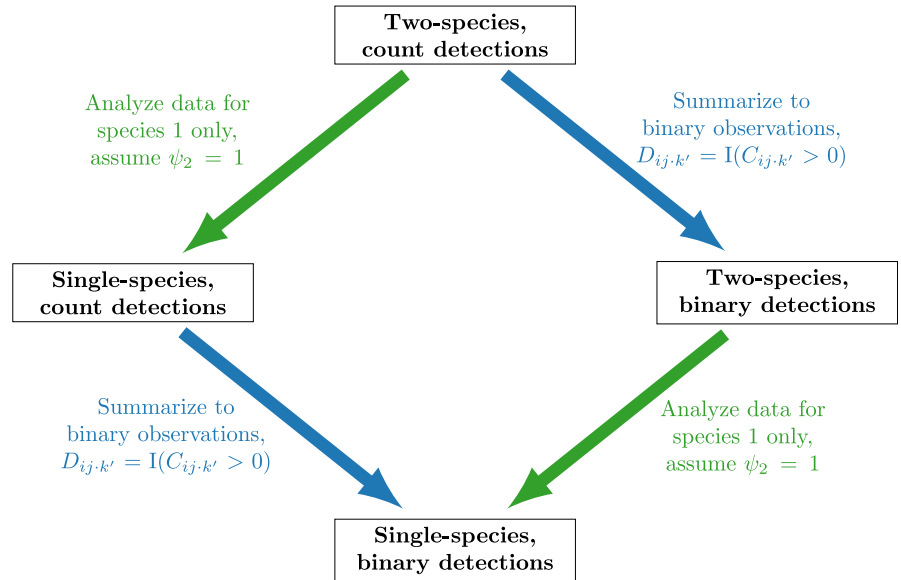
al., 2015). Both single-species models implicitly assume false-positive detections are possible at every site (corresponding to $\psi_2 = 1$ in our example scenario; Figure 2). The assumption that false-positives result from an omnibus source is reasonable in some cases (see Chambert et al., 2015; Chambert, Waddle, et al., 2018), but not for bat acoustic data. Additionally, analysing binary observations instead of detection counts ignores available data.

2.4 | Simulation study

We simulated data for the two-species scenario ($K = 2$) to assess estimates of occupancy and relative activity using different false-positive occupancy models. We simulated datasets assuming our general model (Section 2.2 and Supporting Information S2.1) with scenarios for different data-generating parameter values. Across all scenarios, we assumed species 1 had data-generating parameter values $\psi_1 = 0.3$, $\lambda_1 = 1$, and $\Theta_1 = [0.8, 0.2]$. The data-generating parameter values for species 2 varied across scenarios and included all combinations of $\psi_2 = \{0.25, 0.5, 0.75, 1\}$ and $\lambda_2 = \{0.5, 1, 2, 4\}$. We specified $\Theta_2 = [0.35, 0.65]$ for every scenario, meaning detections from species 2 were misclassified as species 1 with probability 0.35. We simulated 250 datasets consisting of $N = 200$ sites with $J = 4$ visits to each for every scenario. To inform the species classification probabilities, each dataset had two confirmed visits from 50 sites that included unambiguous detection counts for both species. Every dataset was analysed with the data-generating model (our approach) and the three alternative models (Section 2.3 and Supporting Information S2). For the single-species models, we considered species 1 the focal species and consequently only analysed detections classified as species 1. These included the correctly classified detections of species 1 and misclassified detections of species 2. In the models using binary observations, detection counts were summarized to indicator variables (Supporting Information S2).

We focus on the estimates of occupancy and relative activity for species 1 when summarizing the simulation results. For models using binary observations, the relative activity rate is not directly estimated. This parameter can be derived using these models based on the probability of observing a count greater than zero from the Poisson distribution (Supporting Information S2). Additionally, the single-species models are unable to separately estimate the relative activity rate (λ_1) from the probability of correct classification (θ_{11}) and activity estimates from these models correspond to the product of these parameters (Supporting Information S2). For every model and scenario, we plot the average posterior means and average 95% posterior intervals (PIs) associated with occupancy (ψ_1) and relative activity (λ_1) for species 1. The average 95% PIs consist of the average 2.5% and 97.5% posterior quantiles across all simulated datasets. Coverage of the 95% PIs quantifies how often these intervals included the data-generating values ($\psi_1 = 0.3$, $\lambda_1 = 1$).

Models were fit using Bayesian methods with Uniform(0, 1) prior distributions for all probability parameters. In the count models, Gamma(2, 0.25) prior distributions were used for rate parameters (λ). The parameterizations of these models result in different prior distributions, but all distributions we used are sufficiently vague that they have little impact on the resulting posterior inferences. Data were generated in R (version 3.5.1; R Core Team, 2018) and models were fit with stan (Carpenter et al., 2017) using the rstan package (version 2.18.2; Stan Development Team, 2018). The results from a fitted model were excluded if it had algorithm warnings (divergent transitions or reached the maximum tree depth), failed to converge, or had a small effective sample size for any parameter. Each model was fit with four independent chains of 800 total iterations (the first half of each chain was discarded as burn-in). We used the ggplot2 package (version 3.1.0; Wickham, 2016) to create figures. Code used to conduct these simulations is provided in Supporting Information Data S1.

2.5 | Application to bat data

We demonstrate our model using bat acoustic data collected by the Montana Natural Heritage Program and their partners in May 2014

(Maxell, 2015). We provide information on the survey protocols here, but additional details are available in Maxell (2015). Acoustic detectors (Song Meter SM2BAT + ultrasonic records, Wildlife Acoustics, Inc; http://www.wildlifeacoustics.com) were deployed to identify bat species and monitor activity at 40 sites throughout Montana. Site locations were based on local objectives and not a probabilistic sampling design. These sites were part of a long-term acoustic monitoring effort across Montana, but our example analysis focused on a subset of the available data. Within this time period (May 2014), we selected only six nights. Most sites had a detector deployed for all six nights, but some sites had data for fewer nights. We selected nights from throughout the month to reduce potential autocorrelation among consecutive nights. For this analysis, we considered call files summarized for a single night as one visit to a site.

All recorded bat call files were classified using Sonobat (version 4, https://sonobat.com). We analysed data for seven species: big brown bat *Eptesicus fuscus* (EPFU), silver-haired bat *Lasionycteris noctivagans* (LANO), hoary bat *Lasiurus cinereus* (LACI), California myotis *Myotis californicus* (MYCA), western small-footed myotis *Myotis ciliolabrum* (MYCI), long-eared myotis *Myotis evotis* (MYEV) and little brown bat *Myotis lucifugus* (MYLU). We included an 'other' category for additional species occurring in Montana that are difficult to detect acoustically and/or not widespread. We did not believe there was enough data to reliably estimate parameters for each species in the 'other' group. We informed the automated species classification probabilities using auxiliary data from manual identifications for a portion of bat calls recorded during the long-term acoustic monitoring across Montana from 2011 to 2017. The criteria for definitively determining species in call files can be found in Bachen et al. (2018). For bat acoustic data, one challenge with informing the classification probabilities is that most recorded call files cannot be definitively identified manually. Consequently, while the manual species identifications are considered error-free, some detections remain ambiguous after manual examination. Using these auxiliary data assumes that the automated species classification probabilities for call files capable of being manually confirmed do not differ from those that cannot receive manual confirmation. The same assumption is required to use manually confirmed calls from a subset of the observed data and our model can be modified when detections cannot be confirmed (Supporting Information S3).

We included site-level and visit-level covariates to explain heterogeneity in the occupancy probabilities and activity rates for each species. Using the logit link, we modelled the occupancy probabilities with site elevation and an indicator for whether the location was near a lentic (stagnant) water source or not. We used the log link to associate activity rates for each species with the same site-level covariates in addition to air temperature, wind speed, and relative humidity averages each night (obtained from MesoWest, see Maxell, 2015). All covariates were centred and standardized before inclusion in the analysis. We suspected the nightly call counts would be more variable than expected for the Poisson distribution, even after accounting for heterogeneity with the available covariates. To account for potential overdispersion, we added an additional error term to

the estimated nightly activity rate for each species. Specifically, we let $\log(\lambda_{ijk}) = \mathbf{V}_{ij}\boldsymbol{\beta}_k + \varepsilon_{ijk}$ with $\varepsilon_{ijk} \sim \text{Normal}(0, \sigma_k^2)$. The negative binomial distribution also allows for overdispersion in detection counts (e.g. Chambert, Waddle, et al., 2018), but would preclude some computational benefits of the Poisson distribution (see Supporting Information S1).

For each covariate, the corresponding coefficients were modelled with a hierarchical structure across species (e.g. Dorazio & Royle, 2005). For instance, the occupancy intercepts (α_{0k}) were given a normal prior distribution ($\alpha_{0k} \sim \text{Normal}(\mu_{\alpha_0}, \sigma_{\alpha_0}^2)$), allowing information to be borrowed across species. We assumed Normal(0, 9) and Gamma(2, 1) hyper-prior distributions on these means and standard deviations, respectively. We used Half-Normal(0, 9) prior distributions for the overdispersion standard deviations (σ_k). We placed weakly informative Dirichlet prior distributions on the vectors of species classification probabilities ($\boldsymbol{\Theta}_k$). In this case, we specified $\boldsymbol{\Theta}_k \sim \text{Dirichlet}(\boldsymbol{\gamma}_k)$ where each $\boldsymbol{\gamma}_k$ is a K-length vector with the kth element equal to twenty and remaining elements equal to two. These prior distributions place more prior density on probabilities associated with correct species classifications, reflecting an assumption that the software will most likely classify call files correctly. Using Stan (Carpenter et al., 2017) and the `rstan` package (version 2.18.2; Stan Development Team, 2018), we fit this model with four chains of 1,000 iterations and discarded the first half of each chain as burn-in.

3 | RESULTS

3.1 | Simulation study

Across all simulated scenarios, both two-species models resulted in unbiased posterior means and 95% PIs with high coverage for ψ_1 (Figure 3a; squares and triangles for counts and binary observations, respectively). Results were similar for these models, but the average 95% PIs were 5%–14% wider using binary detections compared to using counts (Supporting Information S4). For all scenarios where data were generated with $\psi_2 < 1$, both single-species models resulted in biased posterior means and 95% PIs with lower coverage for ψ_1 (Figure 3a; circles and diamonds for counts and binary observations, respectively). For these models, bias increased and coverage decreased as data were generated with larger detection rates for species 2 (λ_2). Similar patterns were seen for estimating λ_1 – single-species models were biased and 95% PIs had low coverage (Figure 3b). Note that the single-species models can only estimate the product of λ_1 and θ_{11} (see Supporting Information S2). In our simulations, this product was 0.8, but the average PIs for the single-species models increasingly deviated from this value as λ_2 increased. Both the two-species models resulted in unbiased posterior means for λ_1 but the average 95% PIs were 28%–41% wider when analysing binary observations instead of call counts (Supporting Information S4).

Of the scenarios explored, the single-species models had the poorest performance when species 2 was expected to occupy half of the sites ($\psi_2 = 0.5$). When species 2 occurred at every site,

FIGURE 3 From the simulation study, species 1 average 95% PIs for (a) occupancy (ψ_1) and (b) rate of relative activity (λ_1) from four models. The two-species, count detections model is described in the main text. Other models were previously developed – single-species, count detections: Chambert, Waddle, et al. (2018); two-species, binary detections: Chambert, Campbell Grant, et al. (2018); and single-species, binary detections: Chambert et al. (2015). Scenarios differed based on the data-generating values of occupancy probabilities ($\psi_2 = \{0.25, 0.5, 0.75, 1\}$) and activity rates ($\lambda_2 = \{0.5, 1, 2, 4\}$) for species 2. Lines span the average 95% PIs and points indicate the average posterior means. Colours indicate the proportion of fitted models where the 95% PI included the data-generating values for $\psi_1 = 0.3$ and $\lambda_1 = 1$
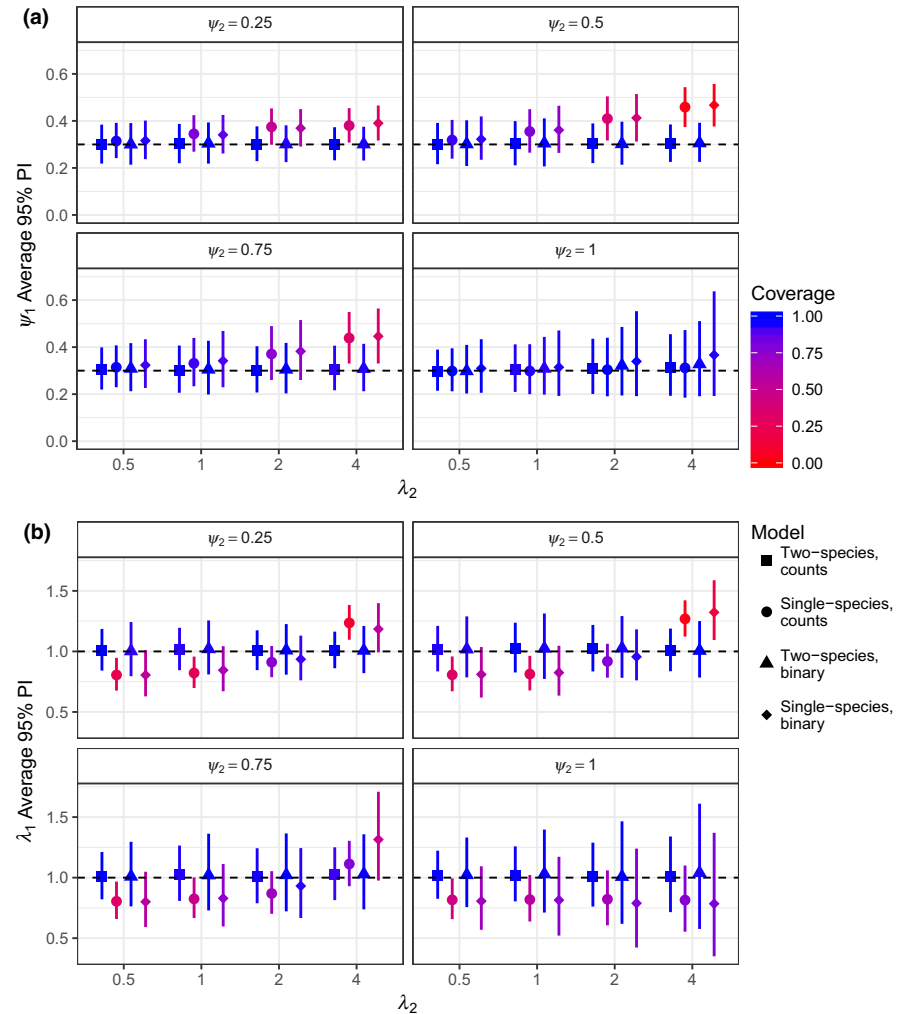
satisfying the assumption implicit in the single-species models, inferences about occupancy for species 1 were improved. Specifically, bias of the posterior means was reduced (although still evident for the single-species, binary detections model) and the 95% PIs had high coverage for both occupancy and relative activity (Figure 3). Across all models, there was more uncertainty for ψ_1 and λ_1 as ψ_2 increased.

3.2 | Bat data example

The estimated probabilities of occupancy varied across the eight bat species explored in this analysis (Figure 4). Some bat species (e.g. big brown bat – EPFU) were estimated to be present at most sites, whereas others (e.g. California myotis – MYCA) were more rare. These patterns are consistent with what is known about the distributions of these species in Montana (Maxell, 2015). Across all species, the probabilities of occupancy were positively associated with lentic water sources and negatively associated with higher elevations, but the 95% PIs for these coefficients generally included zero (Figure 4). All species-specific estimates of relative activity were positively associated with lentic water sources and warmer nights (Figure 5), similar to findings from other studies (e.g. Adams & Thibault, 2006; Parsons, Jones, & Greenaway,

2003). For the other covariates, the estimated relationships with relative activity were more variable across species. There was evidence of overdispersion in the nightly call counts for each species and the posterior distributions for the corresponding standard deviations (σ_k) were generally near 2. For all species, the probabilities of correct classification were high (Figure 6), but some species were more easily classified than others. For instance, call files from MYLU were only correctly classified around half the time and frequently misclassified as species in the 'other' category.

4 | DISCUSSION

We developed a statistical framework for analysing ARU data to estimate occupancy and relative activity for multiple species simultaneously. Our approach addresses some of the challenges posed by multi-species ARU data, including incorporating the species classification probabilities and utilizing the available counts of call recordings. Alternative models fail to capture these characteristics and, consequently, analyses can lead to biased and imprecise inferences. In our simulation study, for instance, two-species scenarios resulted in biased estimates of occupancy and relative
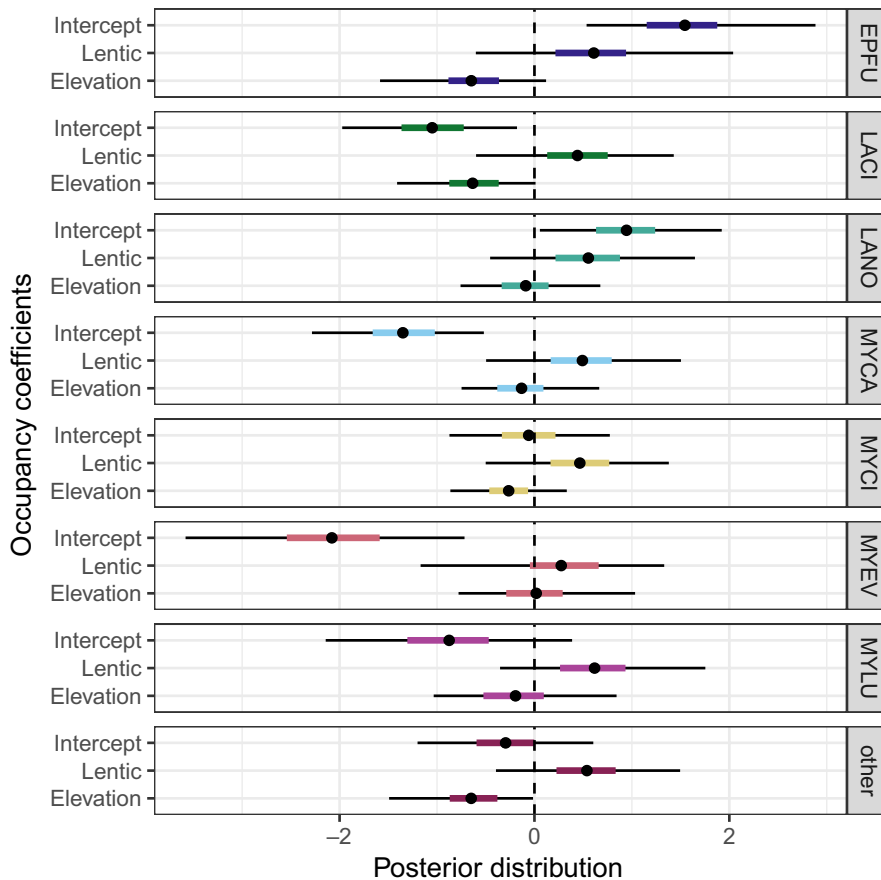
FIGURE 4 From our analysis of the Montana bat data, the occupancy coefficient posterior distributions for each species. Points indicate posterior means. Black lines show 95% PIs and thicker lines show 50% PIs

activity when analysing observations for only one species. Even when including false positives, single-species models had biased estimates because of unmodelled heterogeneity in the probability (or rate) of false positives. Unexplained variability in false positives can bias occupancy estimates in other scenarios as well (Louvrier, Chambert, Marboutin, & Gimenez, 2018). Although various approaches exist for incorporating heterogeneity in false-positive detection probabilities, our model is advantageous when erroneous detections are specifically due to assigning the wrong species to an observation (e.g. recorded call) because it connects the heterogeneity in false positives to the occurrence of other species at a site. Modelling this connection through the species classification probabilities is also important for describing the false-negative detections that result from misclassifying observations. Additionally, simulations demonstrated how utilizing counts of call recordings increases precision when estimating relative activity compared to analysing binary observations. Our model provides a better description of ARU data in order to allow for improved ecological inferences.
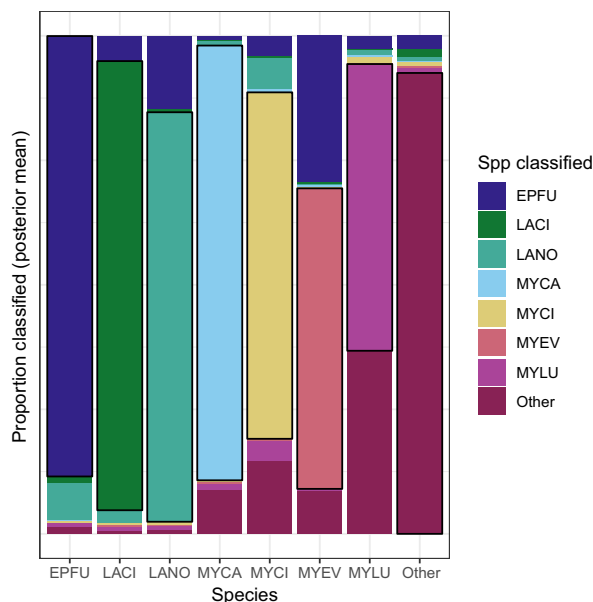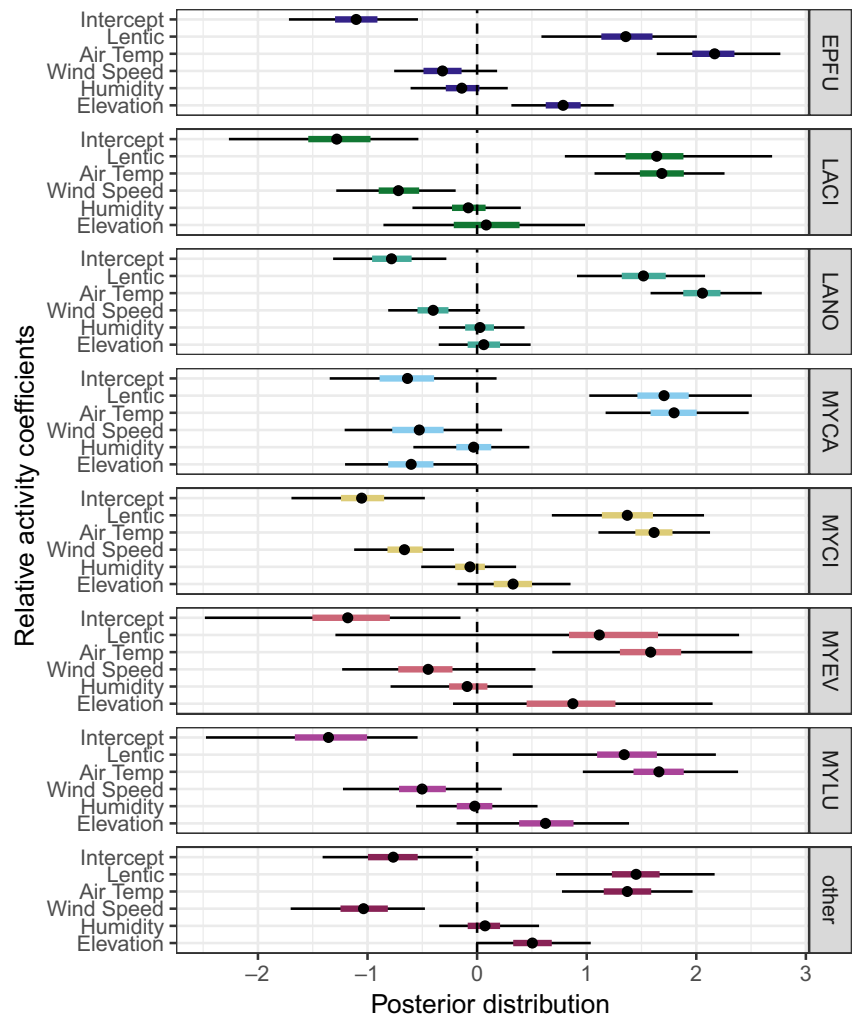
Applying our model to bat acoustic data from Montana allowed for species-specific estimates of occupancy and relative activity. Our analysis incorporated environmental predictors necessary for mapping species distributions and demonstrated that realistic species assemblages ($K > 2$) can be modelled simultaneously. Estimating occupancy and relative activity provides useful inferences for both regional species distribution patterns and local population changes.

Previous work has demonstrated using bat acoustic data for predicting species occurrence over large spatial extents to monitor potential declines (e.g. Rodhouse et al., 2019). Alternatively, estimating relative activity helps address questions at smaller scales, such as how bat populations are affected by disease (Nocera et al., 2019), habitat alteration (Johnson et al., 2008), or artificial light (Straka, Wolf, Gras, Buchholz, & Voigt, 2019). In both cases, species-specific estimates are important to account for species-specific responses (e.g. Nocera et al., 2019). Within a single framework, our approach provides inferences for both occupancy and relative activity while accounting for errors in automated species classifications. By directly modelling the species classification probabilities, we can utilize the automated classifications and reduce the extensive effort typically required to manually confirm species identifications in bat acoustic data.

An important consideration for our model is how to inform the species classification probabilities. The needed information is similar to that for other false-positive occupancy models (Chambert et al., 2015; Guillera-Arroita et al., 2017) or abundance models accounting for species classification errors (Conn et al., 2013). Using an error-free method to classify a portion of the call recordings is beneficial because it reduces uncertainty in occupancy states by confirming species presence at some sites. For bats, manually classifying call files is not only time-consuming and costly, but some recordings cannot be confirmed and their species classifications remain ambiguous. In this case, the confirmed call files need to be representative of the automated classification process to reliably inform these

FIGURE 5 From our analysis of the Montana bat data, the posterior distributions of coefficients for the relative activity rates of each species. Points indicate posterior means. Black lines show 95% PIs and thicker lines show 50% PIs

probabilities. Alternatively, the species classification probabilities can be informed using auxiliary data where the automated process classifies recordings of known species. These recordings must be representative of those collected under field conditions to estimate the species classification probabilities. For bat acoustics, these assumptions deserve more attention in order to collect better data for modelling the species classification probabilities.

In our example, we assumed constant species classification probabilities, but our approach could also be extended to incorporate covariates (similar to Conn et al., 2013). For instance, call characteristics obtained by automated classifiers (e.g. call duration, maximum frequency) could better isolate potential species. This information influences the degree of certainty associated with each classification which has been used in other approaches (Balantic & Donovan, 2019; Barré et al., 2019). General recording quality also influences classification probabilities and could be modelled with environmental conditions during detector deployments. Any potential variability in species classification probabilities would also need to be considered when collecting auxiliary data or confirming species classifications to ensure these data could estimate these relationships. Future work could develop extensions to account for any call recordings that cannot be classified automatically (Conn et al., 2013). Beyond

FIGURE 6 Posterior means for the classification probabilities of call files from each species. Call files from a certain species are shown in each column with the different colours indicating the proportion of calls expected to be classified as each potential species. Black outlines indicate correct classifications

acoustic data, our model could be applied to other survey methods (e.g. cameras) that collect counts of detections where species classification errors are possible.

Acoustic technology is valuable for large-scale monitoring efforts because it efficiently surveys communities for a variety of taxa (Gibb et al., 2019). Collecting ARU data is also beneficial because it can be used to estimate relative activity rates in addition to occupancy. However, fully utilizing ARU data requires modelling the uncertainty in species classifications and obtaining additional information about species classification probabilities. With this information, our model provides a way to efficiently estimate occupancy and relative activity with multi-species ARU data. Our statistical framework helps address the challenges posed by acoustic data, allowing ecologists to better harness this technology when monitoring wildlife communities.

AUTHORS' CONTRIBUTIONS

W.J.W. developed the model, designed the simulation study, conducted statistical analyses, and led manuscript writing. K.M.I. helped develop the model. K.M.I., E.S.A., and A.R.L. helped design the simulation study and interpret results. All authors contributed critically to the drafts and gave final approval for publication.

DATA AVAILABILITY STATEMENT

All data associated with this manuscript is available at https://doi.org/10.5061/dryad.q83bk3jdg (Wright, Irvine, Almberg, & Litt, 2019a). Code for the bat data example is available at https://doi.org/10.5066/P9QK83LD (Wright, Irvine, Almberg, & Litt, 2019b).

ORCID

Wilson J. Wright https://orcid.org/0000-0003-4276-3850

Kathryn M. Irvine https://orcid.org/0000-0002-6426-940X

Andrea R. Litt https://orcid.org/0000-0002-6915-8047

REFERENCES

Adams, R. A., & Thibault, K. M. (2006). Temporal resource partitioning by bats at water holes. *Journal of Zoology*, *270*(3), 466–472. https://doi.org/10.1111/j.1469-7998.2006.00152.x

Bachen, D. A., McEwan, A., Burkholder, B., Hilty, S., Blum, S., & Maxell, B. (2018). Bats of Montana: Identification and natural history. Montana Natural Heritage Program Helena, Montana, USA. Retrieved from http://mtnhp.org/Reports/ZOOMontanaBatID Bachen2018.pdf.

Balantic, C., & Donovan, T. (2019). Dynamic wildlife occupancy models using automated acoustic monitoring data. *Ecological Applications*, *29*(3), e01854. https://doi.org/10.1002/eap.1854

Banner, K. M., Irvine, K. M., Rodhouse, T. J., Wright, W. J., Rodriguez, R. M., & Litt, A. R. (2018). Improving geographically extensive acoustic survey designs for modeling species occurrence with imperfect detection and misidentification. *Ecology and Evolution*, *8*(12), 6144–6156. https://doi.org/10.1002/ece3.4162

Barré, K., Le Viol, I., Julliard, R., Pauwels, J., Newson, S. E., Julien, J.-F., … Bas, Y. (2019). Accounting for automated identification errors in acoustic surveys. *Methods in Ecology and Evolution*, *10*(8), 1171–1188. https://doi.org/10.1111/2041-210X.13198

Carpenter, B., Gelman, A., Hoffman, M., Lee, D., Goodrich, B., Betancourt, M., & Riddell, A. (2017). Stan: A probabilistic programming language. *Journal of Statistical Software*, *76*(1), 1–32.

Chambert, T., Campbell Grant, E. H., Miller, D. A. W., Nichols, J. D., Mulder, K. P., & Brand, A. B. (2018). Two-species occupancy modeling accounting for species misidentification and nondetection. *Methods in Ecology and Evolution*, *9*(6), 1468–1477.

Chambert, T., Miller, D. A. W., & Nichols, J. D. (2015). Modeling false positive detections in species occurrence data under different study designs. *Ecology*, *96*(2), 332–339. https://doi.org/10.1890/14-1507.1

Chambert, T., Waddle, J. H., Miller, D. A. W., Walls, S. C., & Nichols, J. D. (2018). A new framework for analysing automated acoustic species detection data: Occupancy estimation and optimization of recordings post-processing. *Methods in Ecology and Evolution*, *9*(3), 560–570. https://doi.org/10.1111/2041-210X.12910

Conn, P. B., McClintock, B. T., Cameron, M. F., Johnson, D. S., Moreland, E. E., & Boveng, P. L. (2013). Accommodating species identification errors in transect surveys. *Ecology*, *94*(11), 2607–2619. https://doi.org/10.1890/12-2124.1

Dorazio, R. M., & Royle, J. A. (2005). Estimating size and composition of biological communities by modeling the occurrence of species. *Journal of the American Statistical Association*, *100*(470), 389–398. https://doi.org/10.1198/016214505000000015

Gibb, R., Browning, E., Glover-Kapfer, P., & Jones, K. E. (2019). Emerging opportunities and challenges for passive acoustics in ecological assessment and monitoring. *Methodsin Ecology and Evolution*, *10*(2), 169–185. https://doi.org/10.1111/2041-210X.13101

Guillera-Arroita, G., Lahoz-Monfort, J. J., van Rooyen, A. R., Weeks, A. R., & Tingley, R. (2017). Dealing with false-positive and false-negative errors about species occurrence at multiple levels. *Methods in Ecology and Evolution*, *8*(9), 1081–1091. https://doi.org/10.1111/2041-210X.12743

Johnson, J. B., Gates, J. E., & Ford, W. M. (2008). Distribution and activity of bats at local and landscape scales within a rural-urban gradient. *Urban Ecosystems*, *11*(2), 227–242. https://doi.org/10.1007/s11252-008-0055-x

Louvrier, J., Chambert, T., Marboutin, E., & Gimenez, O. (2018). Accounting for misidentification and heterogeneity in occupancy studies using hidden markov models. *Ecological Modelling*, *387*(10), 61–69. https://doi.org/10.1016/j.ecolmodel.2018.09.002

MacKenzie, D. I., Nichols, J. D., Lachman, G. B., Droege, S., Royle, J. A., & Langtimm, C. A. (2002). Estimating site occupancy rates when detection probabilities are less than one. *Ecology*, *83*(8), 2248–2255.

https://doi.org/10.1890/0012-9658(2002)083[2248:ESORW D]2.0.CO;2

Maxell, B. A. (2015). Montana bat and white-nose syndrome surveillance plan and protocols 2012–2016. Montana Natural Heritage Program, Helena, Montana, USA. Retrieved from https://archive.org/details/ZOOWNSPlanAndProtocols20151030/page/n3.

Miller, D. A. W., Bailey, L. L., Campbell Grant, E. H., McClintock, B. T., Weir, L. A., & Simons, T. R. (2015). Performance of species occurrence estimators when basic assumptions are not met: A test using field data where true occupancy status is known. *Methods in Ecology and Evolution*, 6(5), 557–565. https://doi.org/10.1111/2041-210X.12342

Miller, D. A. W., Nichols, J. D., McClintock, B. T., Campbell Grant, E. H., Bailey, L. L., & Weir, L. A. (2011). Improving occupancy estimation when two types of observational error occur: Non-detection and species misidentification. *Ecology*, 92(7), 1422–1428. https://doi.org/10.1890/10-1396.1

Nocera, T., Ford, W. M., Silvis, A., & Dobony, C. A. (2019). Patterns of acoustical activity of bats prior to and 10 years after WNS on Fort Drum Army Installation, New York. *Global Ecology and Conservation*, 18, e00633–https://doi.org/10.1016/j.gecco.2019.e00633

O'Shea, T. J., Cryan, P. M., Hayman, D. T., Plowright, R. K., & Streicker, D. G. (2016). Multiple mortality events in bats: A global review. *Mammal Review*, 46(3), 175–190. https://doi.org/10.1111/mam.12064

Parsons, K. N., Jones, G., & Greenaway, F. (2003). Swarming activity of temperate zone microchiropteran bats: Effects of season, time of night and weather conditions. *Journal of Zoology*, 261(3), 257–264. https://doi.org/10.1017/S0952836903004199

Parsons, S., & Szewczak, J. (2009). Detecting, recording and analysing the vocalisations of bats. In T. H. Kunz & S. Parsons (Eds.), *Ecological and behavioral methods for the study of bats* (2nd ed., pp. 91–111). Baltimore, MD: Johns Hopkins Press.

R Core Team. (2018). *R: A language and environment for statistical computing.* Version 3.5.1. R Foundation for Statistical Computing. Vienna, Austria. Retrieved from https://www.R-project.org/.

Reichert, B., Lausen, C., Loeb, S., Weller, T., Allen, R., Britzke, E., ...Verant, M. (2018). A guide to processing bat acoustic data for the North American Bat Monitoring Program (NABat). Open-File Report 2018-1068, U.S. Geological Survey. Reston, Virginia, USA. Retrieved from https://doi.org/10.3133/ofr20181068

Rodhouse, T. J., Rodriguez, R. M., Banner, K. M., Ormsbee, P. C., Barnett, J., & Irvine, K. M. (2019). Evidence of region-wide bat population decline from long-term monitoring and Bayesian occupancy models

with empirically informed priors. *Ecology and Evolution*, 9(19), 11078–11088. https://doi.org/10.1002/ece3.5612

Royle, J. A. (2004). Modeling abundance index data from anuran calling surveys. *Conservation Biology*, 18(5), 1378–1385. https://doi.org/10.1111/j.1523-1739.2004.00147.x

Royle, J. A., & Link, W. A. (2005). A general class of multinomial mixture models for anuran calling survey data. *Ecology*, 86(9), 2505–2512. https://doi.org/10.1890/04-1802

Stan Development Team. (2018). RSTAN: The R interface to STAN. R package version 2.18.2. Retrieved from http://mc-stan.org/

Straka, T. M., Wolf, M., Gras, P., Buchholz, S., & Voigt, C. C. (2019). Tree cover mediates the effect of artificial light on urban bats. *Frontiers in Ecology and Evolution*, 7, 91. https://doi.org/10.3389/fevo.2019.00091

Wickham, H. (2016). *Ggplot2: Elegant graphics for data analysis*. New York: Springer-Verlag. Retrieved from http://ggplot2.org.

Wright, W., Irvine, K., Almberg, E., & Litt, A. (2019a). Data from: Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity. *Dryad Digital Repository*, https://doi.org/10.5061/dryad.q83bk3jdg

Wright, W. J., Irvine, K. M., Almberg, E. S., & Litt, A. R. (2019b). Code release for bat data analyses in "Modelling misclassification in multi-species acoustic data when estimating occupancy and relative activity": U.S. Geological Survey software release. Retrieved from https://doi.org/10.5066/P9QK83LD

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section.